

## **Taller “Aprendiendo a Trabajar con R”**

- **Modalidad del curso:** Presencial
- **Duración y carga horaria:** La carga horaria del curso es de 15hs. Estas se distribuyen en 5 clases de 3 hs cada una.
- **Síntesis del proyecto**

El presente Taller está basado en el aprendizaje del lenguaje de programación “R”. El mismo aporta a los participantes herramientas prácticas para el manejo de Bases de Datos, análisis estadístico de la misma y presentación de informes.

- **Fundamentación**

En los últimos años se han difundido muchas herramientas estadísticas novedosas para el análisis de información. En particular, el software denominado "R", por tratarse de un software libre, recibe el aporte de investigadores en todo el mundo, y multiplica sistemáticamente sus capacidades.

Principalmente, el programa se destaca por su capacidad de trabajar con grandes volúmenes de información, la posibilidad de utilizar múltiples bases de datos en simultáneo y de realizar gráficos de nivel de publicación. A su vez, la expresividad del lenguaje permite diseñar funciones específicas para las necesidades del usuario.

- **Objetivos**

El objetivo general del curso es que los participantes aprendan la lógica del lenguaje “R”.

De manera más específica se proponen los siguientes objetivos:

- a) Comprender el funcionamiento del programa y la estructura lógica de su lenguaje.

b) Operar sobre las bases de datos: Aplicar filtros sobre la base; seleccionar, crear, renombrar y recodificar variables; unir bases de datos. Crear indicadores de resumen de la información. Realizar gráficos utilizando múltiples variables.

c) Adquirir independencia para continuar el aprendizaje una vez terminado el curso: Conocimientos sobre las “ayudas” del programa, descripciones de los paquetes y funciones; y de los sitios web donde interactúa la comunidad científica usuaria del “R”.

- **Contenidos**

Cada una de las clases del taller estará acompañada por una nota de clase en la cual se plasman conjuntamente las líneas de código que se desarrollarán y la correspondiente explicación de los procedimientos realizados. A su vez los participantes dispondrán de un *script* con el código de cada clase, para ir corriendo secuencialmente cada ejercicio a la par del docente. Finalmente se asignará un espacio de 60 minutos, para realizar un *data clinic*: un espacio para que los participantes traigan datos reales del trabajo sobre los que requieran trabajar con los contenidos vistos en clase, y se discuta y resuelva el problema de forma grupal.

### **Clase 1 - “Conceptos Principales de R Base”:**

- Descripción del programa “R”. Lógica sintáctica del lenguaje R y comandos básicos:

- Características del “R” y Comparación con software similares
- Caracteres especiales en “R”
- Operadores lógicos y aritméticos
- Definición de Objetos: Valores, Vectores y DataFrames
- Lectura y Escritura de Archivos

### **Clase 2 - “Manipulación de datos con Tidyverse”:**

- Funciones principales para el trabajo con bases de datos (paquete “tidyverse”):
  - Renombrar, recodificar y seleccionar variables
  - Funciones para unir diferentes bases de datos.
  - Ordenar y agrupar la base de datos para realizar cálculos
  - Creación de nuevas variables
  - Aplicar filtros sobre la base de datos
  - Construir medidas de resumen de la información

### **Clase 3 - “Gráficos”:**

- Gráficos básicos de R (función “plot”): Comandos para la visualización ágil de la información
- Gráficos elaborados de R (función “ggplot”):
  - Gráficos de línea, barras, Boxplots y distribuciones de densidad.
  - Definición de parámetros de los gráficos: Leyendas, ejes, títulos, notas, colores.
  - Gráficos con múltiples cruces de variables.

### **Clase 4 - “Modelo lineal y análisis estadístico en R”:**

- Correlación lineal:
  - Coeficiente de correlación lineal y test de Pearson
  - Correlograma
- Correlación vs Causalidad
- Teoría del modelo lineal simple:
  - Definición y supuestos del modelo lineal
  - Parámetro, estimador y estimación
  - Mínimos cuadrados ordinarios
- Modelo Lineal Simple usando R:
  - Función “lm”
  - Interpretación de los coeficientes del modelo
- Evaluación del modelo lineal simple:
  - Significatividad de los coeficientes y significatividad global
  - $R^2$  (coeficiente de determinación)

## **Clase 5 - “Shiny y Rmarkdown”:**

- Abordaje de técnicas más sofisticadas en R útiles para automatizar el procesamiento periódico de la información:
  - Loops
  - Creación de funciones a medida del usuario
  - Estructuras condicionales
  - Paquete *Purrr*
- Ayudas del R: Foros reconocidos de usuarios de R, comandos para acceder a la descripción de funciones desconocidas.
- Manejo de las extensiones del software “Rmarkdown” y “RNotebook” para elaborar documentos de trabajo, presentaciones interactivas e informes:
  - Opciones para mostrar u ocultar código en los reportes.
  - Definición de tamaño, títulos y formato con el cual se despliegan los gráficos y tablas en el informe.
  - Caracteres especiales para incluir múltiples recursos en el texto del informe: Links a páginas web, notas al pie, enumeraciones, cambios en el formato de letra (tamaño, negrita, cursiva)
  - Código embebido en el texto para automatización de reportes
- Introducción al diseño de aplicaciones “Shiny”